Crafting Culturally Aligned Narratives: Large Language Models for Arabic Children's Story Generation

Houssam Eddine Boukhalfa Selma Mani Larbi Said Chikh

Intelligent Systems Department | National School of Artificial Intelligence (ENSIA), Algeria

{houssam-eddine.boukhalfa, selma.mani, larbi.saidchikh}@ensia.edu.dz

Motivation

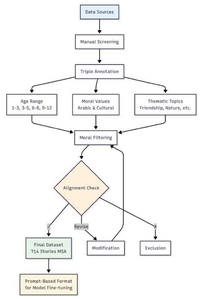
- Traditional Arabic children's storytelling is central for transmitting moral values and Islamic cultural heritage, but high-quality digital Arabic content remains limited.
- General-purpose LLMs underperform for Arabic and often fail to meet age-appropriate and ethical requirements for children
- Need: an LLM-based system that can generate culturally aligned, age-appropriate, and morally explicit Arabic stories

Contribution

- A curated dataset of **714** Modern Standard Arabic stories annotated with age range, topic, and moral values, with a public CC-BY subset of 110 stories.
- Fine-tuning and evaluation of multiple Arabic(-centric) LLMs (**Silma 9B**, **Noon 7B**, **Jais 13B**, **BLOOM 7B**, **Gemini 2.0**) with a focus on cultural and moral alignment.
- A dual evaluation protocol (automatic rule-based + expert human assessment) and a **Story4Kids** mobile app prototype for interactive usage.

Dataset construction

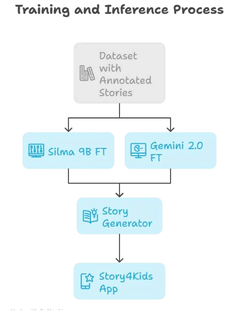
- **714** stories from storytelling websites, educational platforms, and YouTube transcripts, manually screened for narrative coherence, cultural appropriateness, and developmental suitability.
- Each story annotated with:
  - **Age range:** 1–2, 3–5, 6–8, 9–12
  - **Moral value:** honesty, generosity, courage, patience, respect
  - **Thematic topic:** friendship, nature, family, animals
- Critical filtering removed morally inappropriate or culturally misaligned content; inter-annotator agreement: κ = 0.84 (age), κ = 0.79 (moral values).



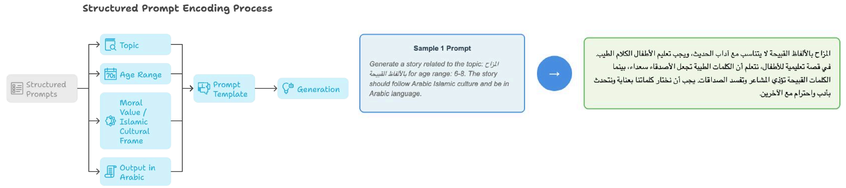
Models and training

Models and fine-tuning

- Evaluated Arabic(-centric) LLMs: Silma 9B, Noon 7B, Jais 13B, BLOOM 7B, Gemini 2.0.
- LoRA fine-tuning (r = 16, α = 32) on RTX A6000 for open models; Gemini 2.0 tuned via Vertex AI.
- Train/val split: **643 / 71** stories (90/10, stratified by age group), 4 epochs, AdamW with cosine LR decay, gradient clipping, weight decay.
- After preliminary experiments, only **Silma 9B FT** and **Gemini 2.0 FT** met cultural and narrative requirements for in-depth evaluation.



Prompting / conditioning



Evaluation and quantitative results

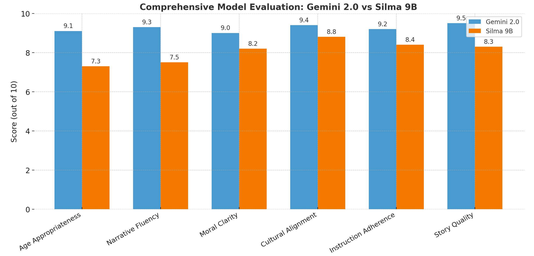
Automatic evaluation metrics

All base→FT improvements are statistically significant ($p < 0.01$).

Model	BLEU	Similarity	Distinct-1	Distinct-2
Silma Original	0.008	0.142	0.08	0.15
Silma Fine-Tuned	0.014	0.197	0.12	0.21
Gemini Original	0.035	0.215	0.11	0.22
Gemini Fine-Tuned	0.051	0.310	0.18	0.34

Human evaluation

- 5 North African Arabic-speaking educators / child specialists (6–15 years experience) evaluated 20 stories per model across 4 age brackets.
- High inter-rater reliability (**ICC = 0.847, Fleiss' κ = 0.78**).
- Gemini 2.0 FT: **9.4/10** cultural alignment, **9.0/10** moral clarity, statistically better than Silma FT (**$p < 0.001$**, large effect size)



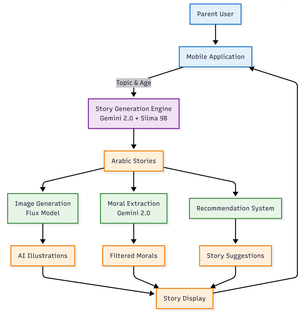
Story4Kids app and qualitative behavior

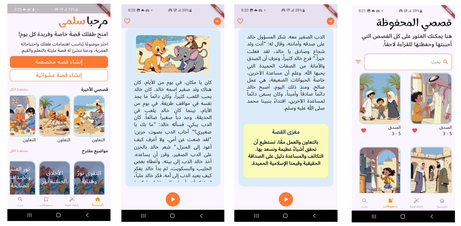
System Design

Story4Kids is an AI-powered mobile application that generates culturally and morally aligned Arabic children's stories. It integrates our fine-tuned LLMs with illustration and moral-analysis modules to deliver age-appropriate, interactive storytelling.

System Components

- Story Generation Engine
- Illustration Module
- Moral Extraction
- User Interface
- Recommendation System





Workflow

- User selects age range & topic
- Story is generated and segmented
- Illustrations are created
- Moral is extracted
- Story + visuals displayed in the app
- Recommendations delivered

References

- Arwa Alhussain and Aqil Azmi. Beyond event-centric narratives: Advancing arabic story generationwith large language models and beam search. Mathematics, 12:1548, 05 2024.
- Silma AI. Silma-9b-instruct-v1.0. https://huggingface.co/silma-ai/SILMA-9B-Instruct-v1.0,2024. Accessed July, 2025.
- Google DeepMind. Gemini 2.0 by google deepmind. https://blog.google/technology/google-deepmind/google-gemini-ai-update december-2024/, 2024.